

Dental Examinations for Quality Control: Peer Review versus Self-Assessment

PETER MILGROM, DDS, PHILIP WEINSTEIN, PhD, PETER RATENER, MS,
WILLIAM A. READ, PhD, AND KENNETH MORRISON, DDS

Abstract: This paper reports the dental care norms for restorative dentistry collected from examinations of 1,466 patients in 105 dental offices in Washington State during 1976. These results are part of a larger study, "Assessment of Care and Continuing Dental Education," being conducted by the University of Washington with the endorsement and cooperation of the Washington State Dental Association. Treatments in volunteer offices were evaluated either by colleagues (peer review) or by the practitioner himself (self-assessment). Two hundred twenty-four of 1,196

eligible dentists volunteered for the study. Patients from the practitioners' offices were randomly selected from office files by project staff. The study tests the proposition that, using standardized clinical evaluation procedures and comparable samples of treatment, dentists will be more critical of their own work than that of others. Results suggest a generally high level of care provided by volunteer practitioners and that self-assessments were significantly more critical than peer review. (Am. J. Public Health 68:394-401, 1978)

Introduction

Increasing attention today is directed toward the quality of health care. This is an area of dentistry which is relatively unexplored: norms are unknown and mechanisms to assure quality are either unevaluated or, more often, untried.

The value, however, of self-examination and peer review is supported by a significant segment of the profession.¹ Interest in such procedures was reported as early as 1846.² However, it was not until 100 years later that a number of public programs, caring for veterans after World War II and for low-income children in Philadelphia and Cleveland, required periodic review.³⁻⁵

Much of the developmental efforts in dentistry quality assurance were developed by Schonfeld and others^{6,7} and Friedman.⁸ The content of good care was defined and methodology developed for post-treatment audit. Experienced practitioners and researchers alike, however, acknowledge the difficulties of quality assurance. Bailit and others^{9,10} have more recently published an analysis of standards devised by members of the Hartford Dental Society and later adopted by the Quality of Care Committee of the Connecticut State Dental Association. After a pretest of the criteria, five dentists were trained to use them in the evaluation of patients. Two independent assessments of 47 dental pa-

tients were made in an effort to determine the reliability of specific items and the practicality of the entire system. The importance of the Connecticut study¹⁰ is that the criteria were developed by practitioners and appear acceptable to dentists in practice. Three components of care were studied: history and examination, treatment planning, and treatment. Treatment criteria for restorative dentistry were among the most reliable measures. Variability was sufficiently high to differentiate between practitioners.

Purpose

This paper reports results of a project in which restorative dentistry was evaluated either by colleagues (peer review) or by the practitioner himself (self-assessment). These results are part of a larger study, the "Assessment of Care and Continuing Dental Education," being conducted by the Department of Community Dentistry at the University of Washington School of Dentistry with the cooperation of the Washington State Dental Association.

The project, begun July 1, 1975, compares the effects of peer review and self-assessment on improving the quality of restorative dental operations. It examines methods of communicating peer review and continuing education information to practitioners. The study also looks at the factors in dental practice that affect the quality of care provided.

In this study, patients were recalled and restorations were examined clinically. Only operative and crown and bridge treatments were assessed. These procedures are well understood by the profession and standards generally accepted. The work by Bailit and others¹⁰ in the development

From the University of Washington School of Dentistry. Address reprint requests to Dr. Peter Milgrom, Assistant Professor and Chairman, Department of Community Dentistry, University of Washington School of Dentistry, Seattle, WA 98195. This paper, submitted to the Journal May 9, 1977, was revised and accepted for publication August 9, 1977.

of standards provided the basis for the design of evaluation criteria used to review treatment in this study.

This experimental study provides restorative dentistry norms and compares peer review and self-assessment. It tests the hypothesis that, using standardized clinical evaluation procedures and comparable samples of treatment, dentists will be more critical of their own work than that of others.

Materials and Methods

Selection of Dentist Subjects

General dentists working more than 25 hours per week in Washington State private practices were eligible for participation in this study. Solicitation was by mail, using a brochure explaining the study, a letter from the president of the dental association encouraging participation, and a postcard to be used to volunteer or request more information. Of 1,196 dentists eligible for the study, 147 dentists returned the postcard marked "volunteer", and 108 dentists requested additional information, 77 of whom eventually volunteered.

Preselection procedures were employed to control sample bias. In order for the participant group to be as representative of the general dental population as possible, the average year of graduation and distribution of dentists in urban King County (Seattle and environs) and the rest of the state (which is primarily rural) were calculated. Each volunteer was assigned a code for one of the four types defined by these two variables. (King County—recent graduate; King County—older graduate; non-King County—recent graduate; non-King County—older graduate.) Using a table of random numbers, participants were selected from each of the four types so that each study group (peer review or self-assessment) would have roughly the same proportion of each type as the general practitioner population.

Evaluation Criteria

The criteria for evaluation of restorative treatment were developed by the senior author (PM), the Chairman of the Department of Restorative Dentistry (KM), and three practicing dentists who were later to perform the actual peer reviews. The written criteria were organized into a series of from 20 to 42 characteristics for amalgams, synthetics, inlays, crowns, and bridges.* An evaluation protocol was developed whereby a trained assistant (recorder) read the individual criteria (e.g., "active caries at the margin?") to which the dentist (assessor) responded with "yes", "no", or "not applicable".

Training

Training was provided for dentist-assessors (peer review) and later for participant dentists (self-assessment). Three assessors were chosen from 24 respected clinicians

with more than ten years experience. All applicants were licensed by Washington State and active in dental affairs.

Training for assessors consisted of closely supervised clinical sessions. Assessors' evaluations of the restorative treatment and oral health found in the mouths of volunteers solicited from the dental school clinics were carefully reviewed; differences between dentists that were identified were discussed at length. Training data for restorative care were converted to quality subscores (see dependent measures). Intraclass correlation¹¹ was calculated after each session as a measure of rater agreement. Subscore values (for r_{cc} , the intraclass correlation coefficient) ranged to .89 for amalgams, .87 for synthetics, .83 for inlays, .78 for crowns, and .78 for fixed bridges. Reliability of .85 for the overall performance of assessors was obtained. Oral health measures were practiced until consistency between raters was achieved.

Self-assessors received a carefully designed one-day orientation session which included an assessment of evaluation skills using dentoform models of clinical defects prepared to provide a wide range of quality of restorative operations. After pretesting, self-assessors talked with members of the staff about the use of the evaluation criteria and procedures. Intraclass correlation measures of reliability for subscores on a post-training assessment were calculated for self-assessors. To facilitate testing, self-assessors were separated into two groups examining different dentoforms. The groups achieved reliability coefficients of .86 and .74 for amalgams and .91 and .84 for inlays, onlays and 3/4 crowns. Results were similar for other types of restorations although the small sample prevented calculation of meaningful correlation coefficients.

Selection of Patients

Patients in both peer review and self-assessment offices were selected by project staff who also functioned as clerical recorders. Interest here was to obtain a variety of types of restorations. A selection system whereby points were assigned to each restoration in a given patient was designed to standardize selection as much as possible and to overselect for crown and bridge treatments which we determined in pretests to be of fairly low frequency. In selecting patients, the recorder began at random in the active patient charts of each practice and examined each chart until she came upon patients with sufficient restorative treatment completed by the participant dentist during the calendar year 1975. Only one year of treatment was reviewed to control for the effects of age on the restoration.

Once an appropriate patient for recall was located, the recorder noted the name, address, telephone number, tooth numbers and types of restorations to be examined on a list. When 40 names were collected, she gave one copy of the sheet to the participant's receptionist and asked that she make appointments for the first 16 patients available on the scheduled day. If a patient was unavailable, the receptionist was instructed to proceed to the next name on the list in order. The recorder would then call at a later date and check to see which patients were actually scheduled in order to prepare data recording sheets prior to the study examina-

*Copies of the criteria are available on request from the senior author.

tion. Since the randomization procedure might result in some names the dentist thought represented atypical or even disagreeable patients, we requested a report on each patient from the dentist. In this way we not only minimized rejection of patients but also were able to assess the dentists' perception of the nonclinical factors affecting patient care and roughly estimate the comparability of these characteristics over the experimental group.

Examination Procedures

For peer review offices, examinations were conducted on a pre-arranged date and patients were scheduled at 15 to 30 minute intervals. One operatory was used, and in most cases where the dentist was present, he continued to work with minimal disruption of office routine. Recorders brought all recording forms and identical prepackaged sterilized sets of a dental mirror, explorer, periodontal probe as well as dental floss, articulating paper, gauze square, and patient napkins. The recorder provided each patient with a written and oral statement of the objectives and procedures of the study and gave him an opportunity to ask questions. Written consent consistent with University policy for the protection of human subjects and the American Dental Association Principles of Ethics was obtained in all cases. Written consent by dentists was obtained at the orientation meeting.

Evaluations were conducted by the project staff using the clinical criteria described earlier. Mirror and explorer or probe and a dental light were employed. No radiographs were used.

For self-assessment offices, patient scheduling was less rigid and, in some cases, exams extended, at the convenience of the dentist and his patients, over several months. After patients were selected by the project staff, the dental assistant in the self-assessment office was instructed in the use of the examination protocol and forms. An illustrated participant guide explained the procedures as well. The dentist and his staff were encouraged to read the manual carefully. Self-assessment recall examinations were conducted by the dentist and his staff according to the same protocol used by the peer review team and adapted for this purpose.

All patients in both experimental groups had their oral health status assessed. Measures employed were DMFT, Vermillion and Greene's Simplified Oral Hygiene Index, and the Lilienthal modification¹² of Russell's Periodontal Index.

Dependent Measures

To summarize the large number of observations, dependent measures are expressed as weighted averages of sets or subsets of specific criteria. Each specific criterion was assigned a weight and subscores were created for types of restorations (amalgams, inlays, etc.). Three overall measures were created from weighted averages of the subscores. All scores were transformed to correct for skew and were scaled to make them easily comprehensible. Two of the overall measures are weighted to emphasize broad definition of quality (SCORE, MORSCORE). The remaining measures emphasized factors that could compromise the longevity of the restoration and result in clinical failure (SERVICE SCORE). All three transformed scores have a possible range

of 0 to 100. A detailed discussion of the weightings is contained in Appendix 1. T-tests were used to test for differences between the self- and peer assessments.

Results

Dentist Sample Characteristics

One hundred five participants were selected among 224 volunteers—65 to be in the peer review group and 40 to be in the self-assessment group. After participants were selected they were notified by phone and asked to attend the orientation meeting. If they could not attend, another volunteer of the same type was selected. Preselection procedures resulted in 38 per cent of dentists from King County compared to 43 per cent in the general practitioner population. The average year of graduation for the study sample was 1960 compared to 1957 in the general population. There were no significant differences between dentists assigned to peer and self-assessment groups.

Patient Sample Characteristics

This sample consisted of 986 patients in the peer review offices and 480 patients in the self-assessment offices. The mean number of patients seen was 15.2 for peer and 13.0 for self-assessment offices. Mean number of restorations per office was 74.1 for peer and 73.7 restorations for self-assessment. Table 1 shows the distribution of types of restorations. Simplified oral hygiene indices for debris and calculus were low for both groups and differences between groups for debris were non-significant. Differences in the calculus level were significant at the $p = .005$ level. Calculus indices were slightly higher for the self-assessment group. There were no significant differences between groups in dental experience (DMFT) or in periodontal status. No patients were excluded based upon the report from the dentists.

Dental Care Norms

For both peer and self-assessment groups the vast majority of year-old restorations were highly satisfactory clinically and fulfilled most evaluation criteria at an acceptable level. As an example, the mean characteristics from peer review are given for the several types of restorations.

Amalgams: Nineteen per cent of amalgams had no negative aspects. Ninety-four per cent had smooth (31%) or fairly smooth (63%) margins. One per cent of alloys had working

TABLE 1—Mean Number of Restorations Examined per Office

Type of Restoration	Peer Review	Self-Assessment
Amalgam	45.1	45.5
Inlays, Onlays, $\frac{3}{4}$ Crowns	3.3	2.2
Synthetics	12.6	15.4
Crowns	10.5	9.0
Bridges	2.7	1.7
Total	74.1	73.7

cusps not protected or poor bulk. Two per cent had occlusal pits and fissures not included in outline; four per cent were clearly underextended gingivally or proximally. Fifteen per cent of alloys had gingival flash; one per cent had major overhangs. One per cent was fractured at the isthmus or had recurrent decay. Six per cent had lots of ditching or a large deficiency. Ten per cent of alloys had an open contact, absence of marginal ridge, or occlusal embrasure. Lesser numbers (2%) had poor contour. Easily correctable interferences were detected in 10 per cent of the cases.

Inlays, onlays, and 3/4 crowns: Forty per cent of single gold restorations (other than full crowns) had no flaws identified clinically. Smooth margins on the occlusal and proximal (59%) and on the gingival (84%). Slight catches on the occlusal or proximal (39%) and slight gingival irregularity (16%) were recorded. Two per cent had unworked or badly overpolished margins. Two per cent had an open or cement margin. Four per cent of inlays had inadequate cusp protection. Anatomy and contour presented no problem in these restorations; however, six per cent had open contacts. All restorations were in centric contact and had no severe interferences in closure or in excursive movements. Fifteen per cent had easily correctable interferences.

Synthetics: Fifty per cent of synthetics had no negative characteristics. Margins were intact (94%) and generally smooth (63%). Six per cent showed a deficiency and one per cent had recurrent decay. Gingival flash (8%) and gingival roughness (8%) were found. Open contact was present in 11 per cent of cases; two per cent were severely over-or-under-contoured. Three per cent were discolored or colors mismatched.

Crowns: Thirty-one per cent of crowns had no negative aspects. All margins were clinically closed and occlusals intact. Four per cent had areas of excess or deficiency. Ninety-one per cent had acceptable marginal contour, but some gingival roughness (20%) was present. Nine per cent of crowns had open contacts. Smaller numbers (2%) were over-or-under contoured. Major color deficiencies were present in some porcelain faced crowns (7%), but few problems were found for shape or texture. Fifteen per cent of crowns had correctable interferences; few (1%) had heavy facets.

Bridges: Seventeen per cent of bridges had no negative reports. Most margins were closed (95%); 5 per cent had open or cement margins or a hole in the occlusal surface. Eighty-five per cent of retainers had good marginal contour, 79 per cent had smooth gingival margins; few (2%) had lots of catches. Eight per cent of retainers were clearly underextended and some (3%) needed cusp protection. Almost all retainers (99%) were properly contoured, but 9 per cent again had open contacts. Ten per cent of bridges had minor color deficiencies. Less than 2 per cent of bridges had problems with pontic design or solder joints. Eighteen per cent had easily correctable interferences; 2 per cent had severe interferences.

Dentist Performance Index

Overall weighted indices of restorative treatment quality (SCORE, MORSCORE) ranged from 60 to 88 for SCORE and 58 to 89 for MORSCORE (possible 0-100). Results of weightings of criteria emphasizing reasons for clinical failure (SERVICE SCORE) ranged from 74 to 95 (possible 0-100).

Differences Between Peer Review and Self-Assessment

There were mean differences between self-assessment and peer review in one of the two general overall weighted

TABLE 2—Difference between Normalized Means for Peer Review and Self-Assessment

	Mean	SD	N	t-test(p)
SCORE				
Peer	75.57	5.19	65	0.030
Self	72.94	6.77	37	
MORSCORE				
Peer	74.52	5.67	65	0.116
Self	72.49	7.14	37	
SERVICE SCORE				
Peer	88.15	4.68	65	0.001
Self	84.16	5.25	37	
SUBSCORES				
Amalgams				
Peer	75.57	4.88	65	0.594
Self	74.97	6.20	37	
Synthetics				
Peer	76.91	8.74	64	0.003
Self	71.00	9.99	37	
Inlays, Onlays, ¾ Crowns				
Peer	75.72	9.78	42	0.638
Self	74.46	11.71	25	
Crowns				
Peer	76.38	9.15	63	0.002
Self	69.80	11.33	36	
Bridges				
Peer	76.64	8.08	54	0.004
Self	69.97	11.60	25	

scores (SCORE) for t-test comparisons. Comparisons using a more conservative weighting emphasizing serviceability of restorations (SERVICE SCORE) showed self-assessors again more critical, this time at the $p = .001$ for the t-test (Table #2).

Plots showing the distribution of transformed overall scores (SCORE) are given in Figure 1. Differences in synthetics, crowns, and bridges accounted for most of the divergence between the overall measures (Figure 2). Differences in the expected direction in subscores were found for outline, occlusion, esthetics, and bridge joints. Greatest differences were found in composite evaluations of gold restorations ($p = .008$). (The relative contribution of various criteria are summarized in Figure 2).

Discussion

Norms. The dental care norms reported for this volunteer sample of practitioners show an overall high level of restorative care consistent with the reputation of the region. Selection procedures for dental offices assured us of a generally representative sample stratified by location and years of experience. Considerable bias existed in the volunteer population as a whole and is discussed in detail elsewhere.¹³ In summary, however, volunteers differed from non-volunteers in some situational determinants of volunteering (subject interest, expectation of a favorable evaluation) but not in the personal characteristics that we measured. Preselection procedures were executed to increase the generalizability of the

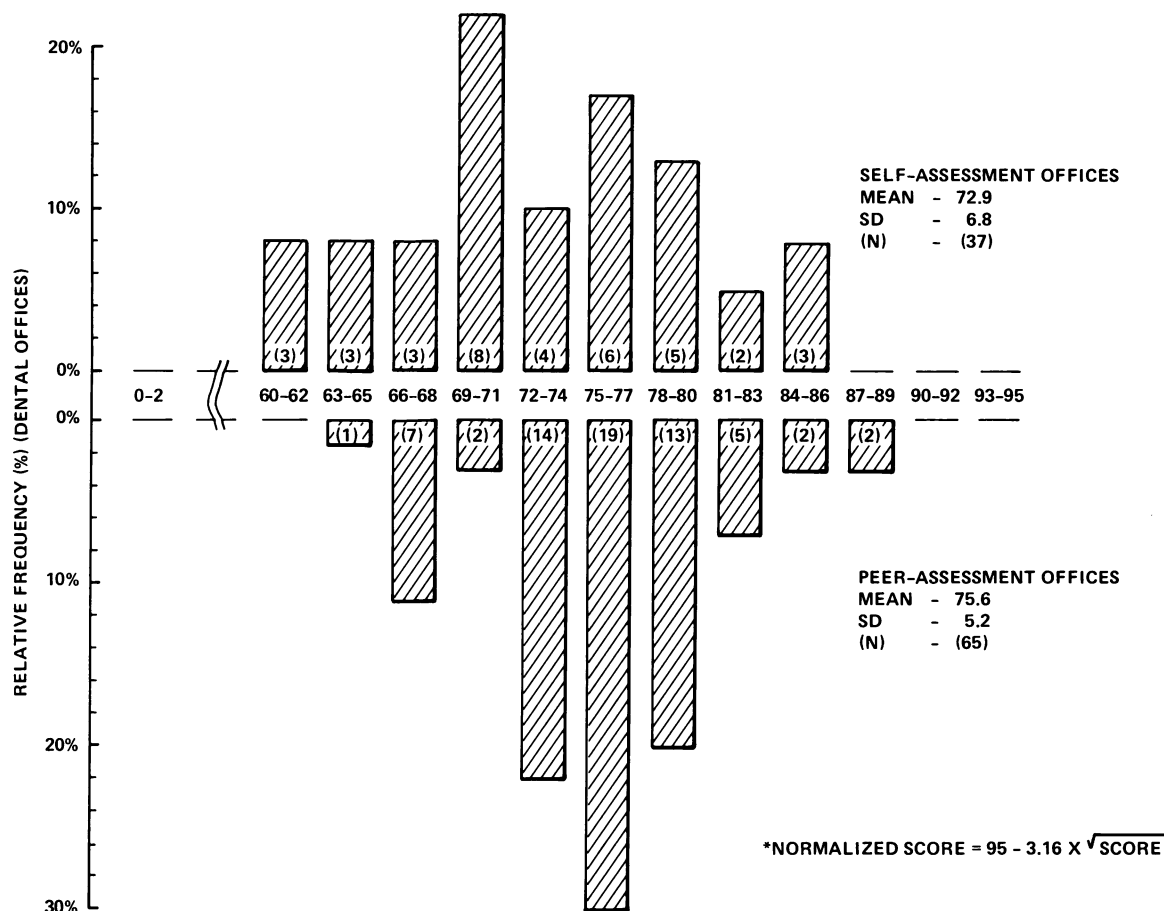


FIGURE 1—Relative Frequency of Normalized Quality Scores for Peer- and Self-Assessment Offices.

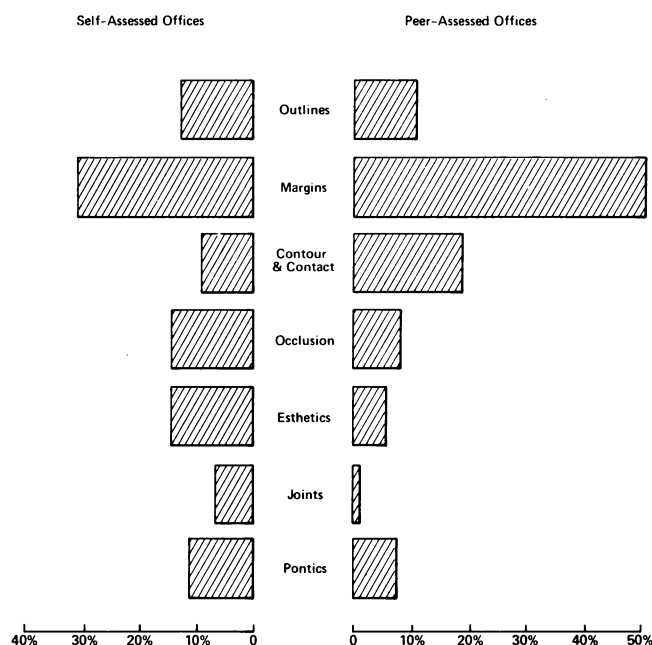


FIGURE 2—Relative Contribution of Chunk Score Categories to Overall Quality Score for Self- and Peer-Assessed Offices.

findings. The norms pose an important question in dental care evaluation. That is, if generally competent dentists do not or cannot meet all evaluation criteria for individual restorations perfectly, what then is an acceptable level of quality for a dental office? Dental schools and licensing boards rightly emphasize the highest level of technical standards for restorations. Nonetheless, dentists should be expected to define the boundaries of satisfactory everyday practice. For example, we uniformly found that about 10 per cent of alloys had an open contact. Is this an acceptable performance standard? One open contact may be clinically unacceptable for the patient. In constructing our quality measures, we found that weighting individual restorative characteristics was helpful in establishing the relative contribution of different defects of individual restorations, but this procedure fails to set a practice-wide standard. No one, to our knowledge, has dealt with this problem. Similarly, no one has identified the relative contribution of factors such as inherent weakness in dental materials or patient oral hygiene that influence clinical outcome but are not under the direct control of one dentist.

Peer review versus self-assessment. Analysis of dependent quality measures confirms our hypothesis that dentists' self-assessments are more critical than in evaluating others, especially in emphasizing those factors that influence ser-

viceability of restorations. There are at least three potential explanations of these differences:

First, the samples of dental offices and patients in each study mode may not be comparable. Dentists were selected and assigned to study groups in such a way as to provide similar distributions of experience and practice location. No other factors—for example, income, dental school attended or practice size—were employed in the stratification procedure. Similarly, patients with recent care were chosen at random from office records. The effects of aging on the serviceability of the restorations were minimized by selecting for examination only restorations placed in the last year. In addition, comparable measures of dental experience (DMFT), oral hygiene (OHI-S), and periodontal status (PI) were taken. Again, differences between groups were fairly well controlled. Nonetheless, factors such as dental IQ, ability or willingness to afford treatment, or complicating medical conditions were not compared. As shown in Figure 2, synthetics and crown and bridge accounted for major differences in overall quality scores. Differences were also noted in outline, occlusion, esthetics, and bridge joints.

Second, differences could be as a result of the assessment procedures. While training of assessors and participants (self-assessment) appears adequate and assessments were apparently carried out according to the protocol, we have no internal comparison of peer and self-assessors evaluating the same set of restorations. Post-test measures of some of the same restorations seen by project staff and self-assessors will be conducted during the next year allowing a check on the reliability of the evaluation process. Our feeling is, however, that t-test statistics are fairly sensitive to measurement variability and that increased reliability would have allowed us to find more differences had they been really present.

Finally, as we have hypothesized, discrepancies between peer review and self-assessment may be a result of actual differences in how critical dentists were in assessing themselves and others. There are no previous controlled experimental comparisons of this nature in dentistry. However, our results are similar to those of Bailit and others¹⁰ in that greater variability existed in the quality of cast gold restorations than for amalgams. Thus, we were better able to detect differences between study groups. Clinical differences for all types of restorations followed a similar pattern although not all were statistically significant. Work by one of us (PW) compared student self-assessment with faculty evaluations of restorative treatment in the dental school clinic.* As students' evaluations were significantly less critical than those of faculty, we believe students to be less sensitive and observant of problems. Other work by Linn, Arostegui, and others¹⁴ in the development of a physician self-performance rating scale suggests that medical students consistently rated themselves lower than they were rated by peers. However,

these ratings were generally not technical or sub-specialty or operation specific. In the case of practitioners, it appears they are more aware of the problems they have encountered in treating the patient and perhaps for this reason are more critical than colleagues. Similarly, since replacement of defective alloys and other restorations is a major activity in many practices,¹⁵ it is possible that on recall practitioners are especially sensitive to elements that affect the longevity of restorations.

These preliminary findings indicate that self-assessment as well as peer review procedures may be useful in conjunction with regular recall programs. As recall schedules receive greater attention as more and more dentists adopt preventive approaches and financial access to care through prepayment increases, reporting or record-keeping mechanisms can be devised through which individual private practitioners could collect self-assessment data on recall. Such information could be reported as part of a university-based continuing dental education or dental society professional development program. Our training data and previous work by Bailit¹⁰ suggest that dentists can be trained to use evaluation criteria in the review of patient care. Mechanisms for developing norms on scores and for processing such data are available through most dental schools.

Conclusion

This study has been directed to the establishment of dental care norms and to the investigation of the utility of clinical peer review and self-assessment procedures. While another round of recall examinations is to be performed in 1977, our findings suggest a generally high level of care provided by our sample volunteer practitioners. More importantly, our review of evaluation procedures confirms the notion that the dentists we studied were quite critical of themselves and the treatments they provided to their patients. Self-assessment, then, may prove to be an effective, useful, and economic procedure in dental quality assurance.

REFERENCES

1. Schonfeld, H. K. Dentists, dental societies, and dental review committees. *J. Conn State Dental Assoc.* 44:140-146, July 1970.
2. Schonfeld, H. K. Peer review of quality of dental care. *J. Amer. Dental Assoc.* 79:1376-1382, December 1969.
3. Cons, N. C. Method of post treatment evaluation of the quality of dental care. *J. Public Health Dent.* 31:104-108, Spring 1971.
4. Clark, C. A., Fintz, J. B. et al. Evaluation of a dental care program. *Health Ser. Rep.* 88(8):763-766, October 1973.
5. Soricelli, D. A. Methods of administrative control for the promotion of quality in dental programs. *Am. J. Public Health* 58:1723-1737, September 1968.
6. Schonfeld, H. K., Falk, I. S., et al. The content of good dental care: Methodology in a formulation for clinical standards and audits, and preliminary findings. *Am. J. Public Health* 57:1137-1146, July 1967.
7. Schonfeld, H. K., Falk, I. S., et al. Professional dental standards for content of dental examinations. *J. Amer. Dental Assoc.* 77:870-877, October 1968.
8. Friedman, J. W. *A Guide for the Evaluation of Dental Care.* Los Angeles: University of California School of Public Health, 1972.

*Weinstein, P., Evans, F. W., et al. Student and faculty ratings of four operative dentistry procedures. University of Washington, Departments of Community Dentistry and Restorative Dentistry and the Office of Research in Medical Education, unpublished data, 1975.

9. Bailit, H. (Chairman), et al. Report of the Quality of Care Committee to the House of Delegates, Connecticut State Dental Association, May 21, 1974. Appendix A. Criteria for the Assessment of General Dental Care.
10. Bailit, H., Koslowsky, M., et al. Quality of dental care. 1. Development of standards. J. Amer. Dental Assoc. 89(4):842, October 1974.
11. Guilford, J. P. Fundamental Statistics in Psychology and Education. New York: McGraw-Hill Co., 1973, page 263.
12. Lilienthal, B., Amerena, V., et al. A comparison of a modified periodontal scoring system with Russell's Periodontal Index. Arch. Oral Biol. 9:575-583, 1964.
13. Milgrom, P., Ratener, P., et al. Volunteering for peer review: Differences between volunteers and practitioners at large. J. Dental Res. 56(Special IssueB):B94, June 1977.
14. Linn, B. S., Arostegui, M., et al. Performance rating scale for peer review and self-assessment. British J. Med. Educ. 9:98-101, 1975.
15. Richardson, A. S., Boyd, M. A. Replacement of silver amalgam restorations by 50 dentists during 246 working days. J. Canadian Dent. Assoc. 39(8):556-559, August 1973.

ACKNOWLEDGMENTS

The peer review examinations were conducted by three teams of a dentist and recorder: Joseph Burke, DDS, Seattle; Milton Johnson, DMD, Vancouver; Lawrence Owens, DMD, Spokane; and Mary Grande, Jane Hatley, and Christina Johnson, all of Seattle. Project personnel also include: Mildred Berto, Administrative Assistant; Anne Fields, Program Assistant; Jeanne Robertson, Office Assistant; and Jean Dennis, Scientific Programmer. In addition, the study staff has been assisted by a professional advisory committee representing the Washington State Dental Association: J. Martin Anderson, DDS, Kent; Johnny Johnson, DDS, Seattle; Richard Lindstrom, DMD, Bremerton; Joseph D. McNally, DDS, Tacoma; Phillip Miollis, DDS, Longview; James K. Muller, DDS, Spokane; Gilbert Phelps, DDS, Seattle; A. Lynn Ryan, DMD, Vancouver; Willmar J. Smith, DDS, Seattle; Keith R. Timberlake, DDS, Seattle; Richard Tucker, DDS, Ferndale; George W. Wood, DDS, Walla Walla. This research is supported by Grant No. 1 R18 HSO1645-01A1 from the National Center for Health Services Research. This paper was presented at the Annual Meeting of the International Association for Dental Research, Copenhagen, Denmark, April 3, 1977.

APPENDIX

Calculation of Quality Scores

The quality scores are various weighted averages of sets of subsets of specific criteria developed to summarize performance characteristics. The weighting, omission of items, and transformation procedure determine the interpretation of each measure. This is a mathematical description of the evaluative scores.

As the basic procedure is similar for each measure, the method will be presented here for calculating SCORE, SERVICEABILITY SCORE, MORSCORE, CHUNK SCORES, and SUBSCORES. The juxtaposition will clarify similarities of method and underscore differences in perspective.

The procedure is as follows:

1. Each specific criterion is assigned a weight. The weight is used to take a weighted average to derive a subcategory score.

For example: The following specific criteria were assigned weights of .088, .044, and 0, respectively:

- _____ Major gingival excess
- _____ Small but definite gingival excess
- _____ No gingival excess

Suppose the dentist had three instances of major gingival excess, seven instances of small but definite gingival excess, and 95 instances of no gingival excess (the latter being the preferred outcome). Then the calculation of the subcategory score would go:

$$\text{Subcategory score} = \frac{.088(3) + .044(7) + 0(95)}{3 + 7 + 95} = .0055$$

2. SUBSCORES are the sums of subcategory scores multiplied by 100 for readability. There are five kinds of sub-scores corresponding to the five types of restorations (criteria) in the study—amalgam sub-score, inlay sub-score, etc. Note that in each criterion, the total of the weights of the most unfavorable subcategory response is one, thus permitting comparison between criteria performance. A sub-score is interpretable as 1,000 times the ratio of observed performance and worst possible performance.
3. SCORE, MORSCORE, and SERVICEABILITY SCORES are the weighted averages of the SUBSCORES (see above), using the number of restorations as weights. That is,

$$\frac{N_A \times \text{Amalgam Sbsc.} + \dots + N_B \times \text{Bridge Sbsc.}}{N_A + N_S + N_I + N_C + N_B} \times 1,000 = \text{SCORE}$$

MORSCORE and SERVICEABILITY SCORE are similarly calculated, with a different set of weights for figuring the subcategory scores.

4. CHUNK SCORES are summaries of performance on sets of subcategories, i.e., "chunks," pertaining to different aspects of restorative work—margins, occlusion, etc. As in the SCORE calculation (above) corresponding chunks of each restoration type are averaged with respect to the number of restorations and multiplied by 1,000. In calculating chunk scores, subcategory scores are computed using chunk score weights. The weights add to one in each chunk.
5. To make scores more understandable and, perhaps, more intuitive, scores were transformed to correct for the expected skewness, and scaled so as to have a range of about 40 points centering somewhere around 75-80. The transformed scores are called Normalized SCORE, Normalized SUBSCORES, etc. The actual transformation used is:

$$x^1 = 95 - 3.16 \sqrt{X}$$

Note that a high score corresponds to favorable performance.

I. Score Formula(Subcategory score)_j

$$= \sum_{i=1}^{n_j} f_i \cdot w_i / \sum_{i=1}^{n_j} f_i$$

$$(\text{Subscore})_k = 1000 \times \sum_{j=1}^{n_k} (\text{Subcat score})_j$$

$$\text{Score} = \left[\sum_{k=1}^5 m_k \cdot (\text{Subscore})_k / \sum_{k=1}^5 m_k \right] \times 1000$$

Where n_j = number of specific criteria in subcategory j . n_k = number of subcategories in restoration type k . m_k = number of restorations of type k . f_i = number of occurrences of specific criterion i . w_i = weight of specific criterion i .

MORSCORE and SERVICE SCORES are calculated in the same manner using different weights (w_i).

II. Chunk Score Formula

CHUNKSCORE

$$= \left[\sum_{k=1}^5 m_k \cdot \left[\sum_{j=1}^{n_{ch}} (\text{subcat score})_j \right] / \sum_{k=1}^5 m_k \right] \times 1000$$

Where (subcategory score) – defined in I.

 n_{ch} = number j of categories in chunk c_h **III. Subscore Formula**SUBSCORE = (subscore)_k × 1000, as in I.**IV. Normalized Score**

$$X' = 95 - 3.16 \sqrt{x}$$

Where x = untransformed score X' = normalized score**NYSPHA to Hold Annual Meeting June 18–21 in Albany**

The New York State Public Health Association will hold its 28th Annual Meeting in Albany, at the Empire State Plaza, from June 18 through June 21, 1978. The theme for the meeting is "Where In Health Are We Going?" The program will examine a variety of ways that limited financial resources will impact on the future of health care in New York State and the nation.

Invited state and national speakers in the fields of health, government, and economics include: New York Governor Hugh Carey; DHEW Assistant Secretary for Health Julius Richmond; Princeton University economics professor Uwe Rhinehardt; Rick J. Carlson, author of "End of Medicine"; Dr. Kevin Cahill, Chairman of the New York Health Planning Commission and the governor's special assistant for health; and Dr. Robert Whalen, New York State Health Commissioner.

Plenary session and workshop topics include national health insurance, health care planning issues, and new approaches to health education and promotion.

The pre-registration fee for four-day conference is \$10; on-site registration is \$15. For further information, contact James M. Stewart, Program Chairman, Box 8650, Albany, NY 12208, or telephone 518/439-7451 or 800/342-9816.